Case Study: CI for population mean

Making an Informed Decision: How Much Should I Spend for this House?
One of the biggest purchases we make in our lifetimes is for a home. Questions that we all ask are these:

- How much should I spend for a particular home?
- How many bathrooms are there?
- How long should I expect a home to be on the market?
- What is the cost per square foot?

The purpose of this project is to help you make an informed decision about housing values. This will help to ensure you receive a good deal when purchasing a home.

(a) Go to a real estate Web site such as www.realtor.com or www.zillow.com and enter the particular zip code you are interested in moving to. Randomly select at least 30 homes for sale and record the following information:

- Asking price
- Square footage
- Number of days on the market
- Cost per square foot (asking price divided by square footage)

(b) For each of the variables identified, determine a 95% confidence interval. Interpret the interval.

(c) Now randomly select 30 recently sold homes and determine the percentage discount from the asking price. This is determined by computing asking

$$\frac{asking\ price - closing\ price}{asking\ price}$$

(d) Determine a 95% confidence interval for percentage discount. Interpret the interval.

(e) For the type of house, you are considering (such as a 2400 square foot 3-bedroom/2-bath home), identify at least 20 homes that are for sale in the neighborhood you are considering. Compute a 95% confidence interval for the asking price of this type of home.

(f) Write a report that details how much you should expect to pay for the type of house you are considering.

Case Study: CI for population proportion

1. You will use Heart dataset uploaded on canvas.
The last column of the data is 'EXANG'. It says if a person has heart disease or not. In the beginning, we have a 'Sex' column as well.
a. What is the point estimate of the female that has heart disease?
b. Construct and interpret a CI for the female population proportion that has heart disease.
c. Construct and interpret a CI for the male population proportion that has heart disease.

3. Use the NYC Flight data and find the 95% CI for the mean TOTAL TIME delay by all the airlines.
Check the requirement of a normal model.
Is a normal model appropriate to find the CI?
If not, how do you want to find CI without the normal model?
Use bootstrap model to compute the same CI and comment on which one is better, with or without normal model (if possible).

Here is the description of the NYC_flight data variables:
Variables
year: Year.
month: Month.
day: Day.
dep_time: Departure time, in Eastern time zone.
dep_delay: Departure delay, in minutes.
arr_time: Arrival time, in the local time zone.
arr_delay: Arrival delay, in minutes.
carrier: Carrier, abbreviated.
tailnum: Tail number of the airplane.
flight: Flight number.
origin: Flight origin, airport code.
dest: Flight destination, airport code.
air_time: Time in the air, in minutes.
distance: Distance between the departure and arrival airports, in miles.
hour: Scheduled departure hour.
minute: Scheduled departure minute.